

# Reconciling mass functions with the star-forming main sequence via mergers

Charles L. Steinhardt,<sup>1,2,3★</sup> Dominic Yurk<sup>2,3</sup> and Peter Capak<sup>2,3</sup>

<sup>1</sup>*Dark Cosmology Centre, Niels Bohr Institute, Blegdamsvej 17, DK-2100 København, Denmark*

<sup>2</sup>*California Institute of Technology, MC 105-24, 1200 East California Blvd., Pasadena, CA 91125, USA*

<sup>3</sup>*Infrared Processing and Analysis Center, California Institute of Technology, MC 100-22, 770 South Wilson Ave., Pasadena, CA 91125, USA*

Accepted 2017 February 24. Received 2017 February 24; in original form 2016 December 8

## ABSTRACT

We combine star formation along the ‘main sequence’, quiescence and clustering and merging to produce an empirical model for the evolution of individual galaxies. Main-sequence star formation alone would significantly steepen the stellar mass function towards low redshift, in sharp conflict with observation. However, a combination of star formation and merging produces a consistent result for correct choice of the merger rate function. As a result, we are motivated to propose a model in which hierarchical merging is disconnected from environmentally independent star formation. This model can be tested via correlation functions and would produce new constraints on clustering and merging.

**Key words:** galaxies: evolution – galaxies: star formation.

## 1 INTRODUCTION

A key discovery over the past decade has been the development of the star-forming ‘main sequence’. Almost all star-forming galaxies at any fixed redshift  $z < 6$  are observed to have a tight correlation between their star formation rates (SFR) and their existing stellar masses (cf. Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007; Peng et al. 2010; Tasca et al. 2015). Studies using a variety of selection criteria, SFR and stellar mass ( $M_*$ ) indicators both show strong agreement at a common redshift and are well fitted with a common exponential decline in SFR at different choices of fixed  $M_*$  as a function of time (Speagle et al. 2014). There is now a strong consensus understanding of the rate at which typical star-forming galaxies make new stars at nearly every redshift where star-forming galaxies are observed.

The more carefully the implications of the star-forming main sequence are considered, the more surprising it appears to be. Although it has been robustly measured, it appears to conflict with theoretical expectations and other observations in two key ways:

(i) Lack of environmental dependence or individuality: Star-forming main sequence studies typically cannot resolve individual galaxies or determine their environment. However, colour-based selection of star-forming galaxies should include all types of galaxies, and thus sample a wide range of environments. After all, galaxies are observed to form in environments ranging from the centres of large clusters (Abell, Corwin & Olowin 1989) to field galaxies (van Dokkum 2005) and even in near-voids (Szomoru et al. 1996).

This environment is observed to be an important factor in determining merger rates (Fakhouri & Ma 2009) and AGN activity (Khabiboulline et al. 2014; Satyapal et al. 2014). Numerical simulations also find that environment should be an important driver of star formation (Hirschmann et al. 2014; Genel 2016). The importance of environment is also underscored by a correlation between star formation in central and satellite galaxies, an effect that has been termed ‘galactic conformity’ (Weinmann et al. 2006; Hartley et al. 2015; Kawinwanichakij et al. 2016).

However, the narrowness of the star-forming main sequence indicates that for a star-forming galaxy, the SFR can be determined nearly exclusively by the stellar mass and cosmic epoch, with all other factors having minimal impact. To within the  $\sim 0.2$  dex scatter of the main sequence, it is not necessary to know the environment in order to determine the SFR. Nor is it necessary to know the morphology, metallicity, star-formation history, age of the stellar population, etc. The main sequence instead indicates that even though individual galaxies end up being unique, their star-formation obeys a universal law independent of local conditions.

(ii) Are large or small Galaxies more efficient? The slope of the star-forming main sequence is less than unity, so that more massive galaxies have higher SFR but lower SFR per unit mass (specific star formation rate or sSFR). Thus, more massive galaxies are less efficient at star formation, and would take longer to form their existing stellar mass if it were all formed on the main sequence. Recent work suggests that the SFR– $M_*$  relation may further flatten at high stellar masses, resulting in large galaxies being even less efficient per unit mass than previously believed (Whitaker et al. 2014; Lee et al. 2015; Tomczak et al. 2016).

However, a variety of other observations instead find that more massive galaxies are more efficient. Although more massive haloes

\* E-mail: [charles@dark-cosmology.dk](mailto:charles@dark-cosmology.dk)

virialize later than less massive ones (Press & Schechter 1974; Sheth, Mo & Tormen 2001; Springel et al. 2005; Vogelsberger et al. 2014), the most massive galaxies appear to finish their star formation earlier than less massive ones, an effect often termed mass ‘downsizing’ (cf. Cowie et al. 1996). Further, in the mass regime most commonly probed by the star-forming main sequence ( $\sim 10^{8-10} M_\odot$ ), at higher halo masses a greater fraction of the baryonic mass is processed into stars, with efficiency peaking around a halo mass of  $10^{12} M_\odot$  (Leauthaud et al. 2012b; Gonzalez et al. 2013). Finally, both ULTRAVISTA (Ilbert et al. 2013) and ZFOURGE (Tomczak et al. 2014) find that the low-mass slope of the stellar mass function flattens towards low redshift, with ULTRAVISTA reporting a sharper effect than ZFOURGE. However, if low-mass galaxies grow more quickly than high-mass galaxies, this will instead produce a steepening slope (Peng et al. 2014).

As a result, we are motivated to search for ways to reconcile the main sequence with the evolution of the mass function. In this work, we build an empirical model based upon observed galactic stellar mass distributions over a wide range of redshifts. Where possible, we have picked the simplest possible prescriptions drawn directly from observational results. To some extent, this goes against the trend of recent modelling (Conroy, Gunn & White 2009; Leja et al. 2015), in which the goal has been to build an increasingly detailed picture of the important physics that drive star formation and galaxy evolution. Our work, by contrast, seeks to find a minimal model consistent with the evolution of the low-mass end of the observed stellar mass function. We ultimately produce a model including star formation along the star-forming main sequence, turn-off constrained to match observed quiescent populations and mergers as independent, history-free events.

In Section 2, we demonstrate that, as predicted by Peng et al. (2014), the observed evolution along the star-forming main sequence alone cannot reproduce observed mass functions. In Section 3, we demonstrate that the observed quenching of some galaxies towards low redshift cannot resolve this discrepancy.

Peng et al. (2014) suggested that an appropriate choice of merger rate function might fix the problem, but had insufficient data to determine whether the required merger parameters would be physically reasonable. In Section 4, we show that the correct choice of merger rate function, combined with the star-forming main sequence, can indeed reproduce the observed evolution in galactic mass functions. Further, this merger rate function is also supported by numerical simulations. As a result, we are able to produce a new model, described in Section 5, in which the star-forming main sequence and mergers combine to match observed galaxy distributions over a wide range of redshift. This model also makes specific predictions for clustering and merging parameters that are currently poorly constrained by existing observations.

This work uses a  $(h, \Omega_m, \Omega_\Lambda) = (0.704, 0.272, 0.728)$  cosmology (Planck Collaboration XIII 2016) throughout.

## 2 THE STAR-FORMING MAIN SEQUENCE AND MASS FUNCTIONS

It is now the observational consensus that almost all star-forming galaxies lie on a ‘main sequence’, linking stellar mass, SFR and redshift at least out to  $z = 6$  (Noeske et al. 2007; Duncan et al. 2014; Steinhardt et al. 2014). This relation has a relatively low scatter ( $\sigma \approx 0.2$  dex at all redshifts; Speagle et al. 2014). It would therefore appear that most galaxies at a common mass formed their stars at

a similar rate and time, and thus share a common history of star formation (Steinhardt & Speagle 2014).

In principle, it should be possible to constrain this common history through a continuity analysis, beginning with a measured stellar mass function at some initial cosmic epoch  $\tau_i$  and requiring that the net effects from star formation, aging of the existing stellar population and merging over a given period of time  $\Delta\tau$  combine to produce the measured stellar mass function at some final  $\tau_f$  (cf. Ilbert et al. 2013).

Previous work along these lines (Behroozi, Wechsler & Conroy 2013; Peng et al. 2014; Tomczak et al. 2016; Contini et al. 2017) has resulted in the development of increasingly complex models in order to attempt to match the observed evolution of the mass function. For example, Tomczak et al. (2016) found that star formation and merging alone required an unphysically high merger rate, whereas Contini et al. (2017) also include stellar stripping in an attempt to better match stellar mass functions. The many possible parameters and complex feedback mechanisms involved in galactic evolution result in models very easily becoming underconstrained by observation. Here, we search for the minimal model that is consistent with observed mass functions.

A natural first attempt is to consider whether stellar mass growth in typical galaxies could be dominated by evolution along the star-forming main sequence, with negligible influence from mergers and other environmental factors. After all, at any given time, only a small fraction of high-redshift galaxies are undergoing a major merger (Leauthaud et al. 2012a), and at redshifts  $z > 1$  relatively few galaxies with  $\log(M_*) < 10.5$  are observed to be quiescent (Ilbert et al. 2013). The remainder will lie on the star-forming main sequence during the period when they apparently form most of their stars.

Qualitatively, observed stellar mass functions are characterized by two properties, each of which we must be able to match: (1) they are reasonably approximated by Schechter functions

$$n(M)dM = \phi e^{-M/M^*} (M/M^*)^\alpha \frac{dM}{M}, \quad (1)$$

at all redshifts  $z < 4$  where they have been well measured and (2) the low-mass slope of the Schechter function is initially steep ( $\alpha \sim -1.6$ ), with a much larger number density of galaxies with  $M_*/M_\odot \sim 10^9$  than  $M_*/M_\odot \sim 10^{10}$ , but flattens out gradually towards  $\alpha \sim -1.4$  at lower redshifts ( $z < 1.5$ , Ilbert et al. 2013).

The principal study from which the observed star-forming, quiescent and overall stellar mass functions used in this work are drawn (Ilbert et al. 2013) fits some mass functions with a standard Schechter function and others with a double Schechter function,

$$n(M)dM = e^{-M/M^*} [\phi_1 (M/M^*)^{\alpha_1} + \phi_2 (M/M^*)^{\alpha_2}] \frac{dM}{M}. \quad (2)$$

Because the double Schechter function was adopted primarily in order to produce the bright-end shape and this work is primarily concerned with the faint-end slope, this difference should be negligible. We have chosen to use a single Schechter function at all redshifts for consistency, re-fitting the mass functions in Ilbert et al. (2013) to produce new parameters where required (Table 1). For the remainder of this work, we describe observed mass functions in terms of the three parameters of their best-fitting Schechter functions: a normalization  $\phi$ , turnover mass  $M_*$  and slope  $\alpha$ .

Speagle et al. (2014) find that the slope of the star-forming main sequence is well fit by  $\text{SFR}/M_* \sim M_*^{-0.16-0.026t}$  since at least  $z \sim 4$  (where  $t$  is the age of the Universe in Gyr), indicating that larger galaxies have lower sSFR (sSFR =  $\text{SFR}/M_*$ ). In Peng et al. (2014) it was shown that main-sequence laws of the form  $\text{sSFR} \sim M_*^{-\beta}$

**Table 1.** Best-fitting Schechter function parameters for observed total, star-forming and quiescent mass functions at  $0.2 < z < 4.0$ , adapted from Ilbert et al. (2013).

$z$	$\log(M_*/M_\odot)$	$\phi$ ( $10^{-3} \text{ Mpc}^{-3}$ )	$\alpha$
0.2–0.5	10.88	1.88	−1.25
SF only	10.60	2.14	−1.23
Quiescent	10.91	0.94	−0.95
0.5–0.8	11.03	0.97	−1.35
SF only	10.62	1.52	−1.29
Quiescent	10.93	1.11	−0.46
0.8–1.1	10.87	1.33	−1.32
SF only	10.80	0.82	−1.40
Quiescent	10.81	1.57	−0.11
1.1–1.5	10.71	1.56	−1.27
SF only	10.67	1.31	−1.27
Quiescent	10.72	0.70	0.04
1.5–2.0	10.74	0.86	−1.39
SF only	10.66	0.94	−1.39
Quiescent	10.73	0.22	0.10
2.0–2.5	10.74	0.51	−1.33
SF only	10.73	0.46	−1.35
Quiescent	10.59	0.10	0.88
2.5–3.0	10.76	0.29	−1.43
SF only	10.90	0.19	−1.49
Quiescent	10.27	0.003	3.26
3.0–4.0	10.74	0.12	−1.54
SF only	10.74	0.12	−1.56
Quiescent		Too few	

for some positive constant  $\beta$  produce steepening in the faint end slope of the galactic mass function over time. Their analytical results do not formally hold for a time-varying  $\beta$ , but the fact that  $\beta(t) = -0.16 - 0.026t$  is always negative indicates that a similar overall steepening should occur for evolution along our updated main sequence. Recent evidence that the high-mass end of the star-forming main sequence may be turning over (Lee et al. 2015) would provide a larger  $\beta$  and an even sharper steepening above  $M_*$ , although this work focuses on the low-mass end.

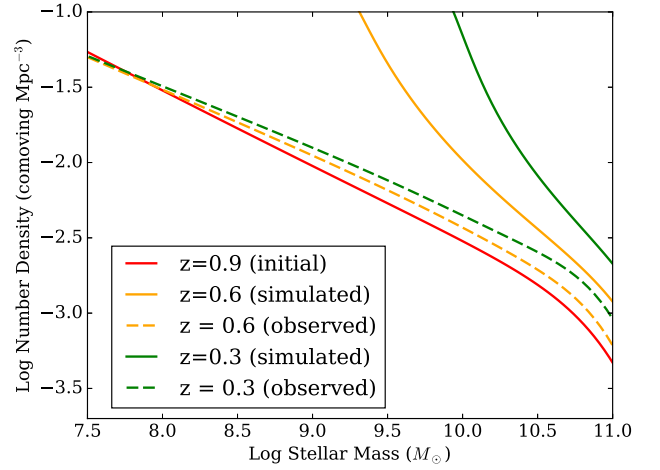
## 2.1 Simulated evolution

In order to investigate this, galaxy populations drawn from higher redshift observed stellar mass functions were evolved along the central values of the star-forming main sequence, then compared with observations at lower redshifts. The goal of these simulations was to develop the simplest model consistent with observed mass functions. Therefore, the initial scenario included only main-sequence star formation, since this is most directly observed. In following sections we also include models for quiescence (Section 3) and mergers (Section 4) in order to produce a better match with observation.

Galaxies drawn from the observed  $z = 0.9$  mass function (Ilbert et al. 2013) were evolved along the star-forming main sequence until  $z = 0.3$ , with the assumption that they remained on the main sequence for that entire time (Fig. 1).

The resulting mass functions have a far different functional form than the observed mass function. If approximated with a Schechter function despite the poor fit, these populations increase in faint-end slope from  $-1.48$  ( $z = 0.9$ ) to  $-2.04$  ( $z = 0.3$ ). However, the observed stellar mass function instead becomes shallower, with a best-fitting slope of  $-1.40$  at  $z = 0.3$ .

Since the exponent in the star-forming main sequence sSFR is always negative, this steepening would be even sharper if the



**Figure 1.** Simulated stellar mass function evolution from  $z = 0.9$  (red) to  $z = 0.3$  (green) along the star-forming main sequence, beginning with the observed stellar mass function at  $z = 0.9$ . This main-sequence evolution produces a sharp increase in the faint end slope of the best-fitting Schechter function, from  $-1.48$  at  $z = 0.9$  to  $-2.04$  at  $z = 0.3$  (solid). Observed mass functions instead have shallower slopes (dashed).

simulation had been started from higher redshifts. However, the faint end of the observed stellar mass function instead becomes slightly flatter towards low redshift. This effect can be reduced if galaxies have a low duty cycle for star formation, but as long as the time-averaged sSFR is lower for more massive galaxies, the stellar mass function will steepen towards low redshift. Moreover, the total stellar mass produced far exceeds that observed at low redshift, so that even if the slope were correct, the number density of galaxies would be too high at all masses. We therefore conclude that, as suggested by Peng et al. (2014), evolution along the main sequence alone cannot account for the observed evolution in the mass function.

## 2.2 Computational limitations

In principle, it would be best to track a large population of galaxies, evolving each independently along the star-forming main sequence. However, the computational complexity and memory requirements for such a study would have severely limited the space of merger models considered in Section 4. As a result, rather than evolving individual galaxies, the simulation instead evolves the mass function as a whole, dividing it into small bins and tracking the number density and its evolution in each bin. The key problem is that mergers between galaxies at very different masses are both common and important to the mass function evolution, so galaxies over a very broad mass range must all be considered as part of the same simulation. However, the number density of galaxies changes by many orders of magnitude over such a mass range. Thus, tracking a statistically significant sample of rare, high-mass galaxies would have required simultaneously tracking too many accompanying low-mass galaxies to be computationally feasible.

Such a strategy requires a more careful consideration of mathematical precision, rounding and binning errors than tracking individual galaxies. In addition to verifying these choices were theoretically unbiased, our results were compared with a simulation tracking smaller numbers of individual galaxies in order to confirm their validity. Each simulation was primarily analysed over  $z = 1.7$  to  $z = 0.3$ , the range for which both the total and quiescent mass

functions are well constrained in Ilbert et al. (2013), with additional but lower quality constraints available from  $z = 2.7$  to  $z = 1.7$ . Because small errors in mass functions build up over the course of the simulation, our results are primarily presented in intervals from  $z = 1.7$  to  $z = 0.9$  or  $z = 0.9$  to  $z = 0.3$ , with both ranges showing similar behaviour.

### 3 QUIESCENCE

Although nearly all galaxies are star forming at the highest redshifts (Steinhardt et al. 2014; Bouwens et al. 2015), at lower redshifts there is an increasingly numerous quiescent population. There are two main ways in which galaxies might appear quiescent; either (1) each galaxy goes through a continuous period of star formation for some length of time and then enters a permanent state of quiescence or (2) each galaxy goes through alternating periods of star formation and quiescence. Both would reduce the total stellar mass produced, albeit in different ways.

#### 3.1 Permanent turn-off

The simplest way to model permanent turn-off is to choose a quiescent mass function  $n_q(M_*, t)$  as a function of cosmic epoch, then at each timestep, remove galaxies from the star-forming population accordingly, freezing their stellar masses. For  $n_q(M_*, t) = 0$ , all galaxies grow continuously along the star-forming main sequence as in Section 2.

Mathematically, it must always be possible to choose  $n_q$  such that the resulting simulated mass functions match observed mass functions. To find such  $n_q$ , after each time step the observed star-forming mass function at the new redshift was interpolated from binned mass functions and the correct number density of galaxies was pulled out of the star-forming population to match that interpolation. All galaxies pulled out of the star-forming population were assumed to have become permanently quiescent.

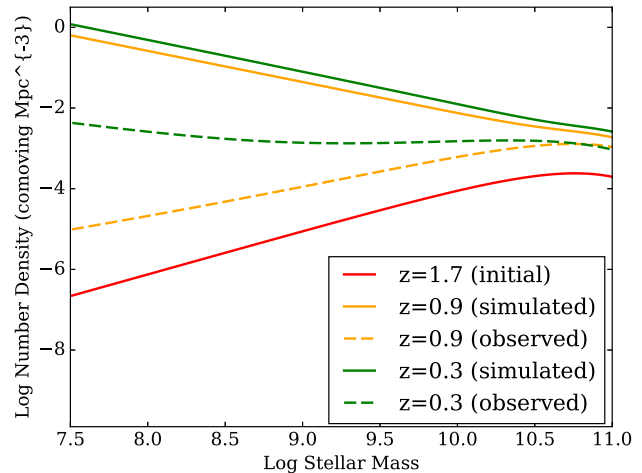
Beginning with the observed star-forming and quiescent mass functions at  $z = 1.7$  (Ilbert et al. 2013), it is therefore possible to find  $n_q(M_*, t)$  such that the total mass functions at  $z = 0.9$  and  $z = 0.3$  are matched by this model (Fig. 2).

Following this path, the predicted quiescent mass functions require a far more numerous population than observed at lower redshifts. This indicates that permanent galactic quiescence does not solve the high-mass overproduction problem created by main-sequence star formation.

#### 3.2 Duty cycles

We now consider whether lower mass galaxies might only become temporarily quiescent rather than permanently turning off, alternating between star-forming and quiescent periods. If star formation has a duty cycle of, e.g. 60 per cent, there would be 60 per cent as much stellar mass growth as implied by the star-forming main sequence. Because quiescence would be temporary, it might be hoped that this could reduce the growth of low-mass galaxies without the permanent turn-off that overproduces quiescent, low-redshift galaxies. Applying a mass-independent duty cycle would not change the exponent in the main sequence SSFR and therefore would not fix the steepening faint-end slope. Thus, in order to match the observed star-forming mass function, the duty cycle must be mass dependent.

It should also be noted that studies of the star-forming main sequence do not measure an instantaneous SFR, but rather attempt to



**Figure 2.** Quiescent mass functions derived from combining main-sequence growth for star-forming galaxies with permanent turn-off, constrained to match observed mass functions from Ilbert et al. (2013) at  $1.7 < z < 0.3$ . The resulting quiescent mass functions  $n_q(M_*, t)$  at  $z = 0.9$  (yellow) and  $z = 0.3$  (green) are compared with observed quiescent mass functions given in Ilbert et al. (2013). Although mathematically the observed total mass function can be matched at all redshifts, doing so predicts a far larger quiescent population than observed at lower redshifts.

estimate the number of luminous, high-mass, blue stars with lifetimes  $\sim 10^7$  yr that dominate the spectral energy distributions of young stellar populations. The main sequence thus describes average star formation rates over  $\sim 10^7$ – $10^8$  yr in star-forming galaxies (Kennicutt & Evans 2012; Hayward et al. 2014). As a result, the periods of quiescence or star formation in this model would need to occur on longer time-scales, and galaxies temporarily in the ‘off’ state when observed would be selected as part of the quiescent population or green valley at that redshift, even though at lower redshifts they might again be star forming.

Following a similar procedure to Section 3.1, our simulation begins with the observed mass function at  $z = 1.7$  and evolves it forward, assigning each galactic mass bin the duty cycle required to match the observed evolution. This would require a duty cycle as low as 10 per cent, so that 90 per cent of star-forming galaxies would actually be selected as quiescent at  $1.7 < z < 0.3$  (Fig. 3), in addition to 100 per cent of post-turn-off galaxies.

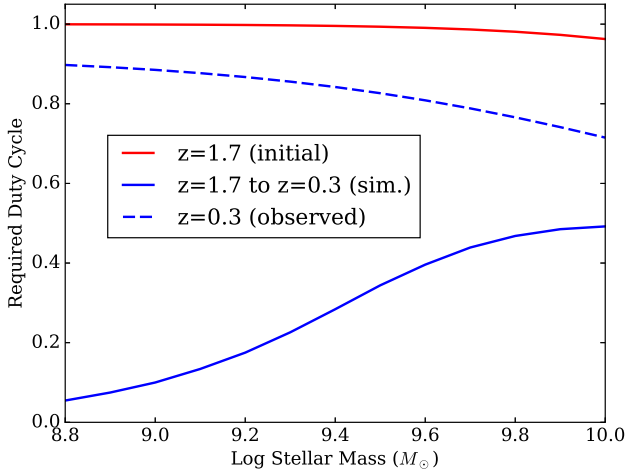
However, at these redshifts, at most 40 per cent of all galaxies with  $\log M_* < 10$  are selected as quiescent (Ilbert et al. 2013). Since neither quiescent scenario can come close to producing the correct evolution in the galactic mass function, the solution must lie elsewhere.

### 4 MERGERS AS A SOLUTION

Finally, we consider the possibility that mergers might flatten the slope of the mass function. This would be a surprising answer because the star-forming main sequence implies that stellar mass growth is similar on cosmic scales, yet mergers are inherently environmental and local. Mergers also would flatten the mass function via a characteristic mechanism: rather than reducing the stellar mass growth in low-mass galaxies, those galaxies grow quickly but then disappear, absorbed into more massive ones.

Although major mergers between two galaxies of similar mass are relatively rare, simulations suggest that the absorption of small





**Figure 3.** Average duty cycle (blue, solid) for the star formation that would be required as a function of mass in order to evolve the mass function from  $z = 1.7$  to  $z = 0.3$  consistent with observations. Low-mass galaxies must have lower duty cycles in order to flatten the slope of the resulting function. Although this quiescence is transitory, these low duty cycles still predict large low-mass quiescent populations at these redshifts. These would be far larger than the observed quiescent fraction of galaxies (Ilbert et al. 2013; red, blue dashed), particularly on the low-mass end.

galaxies by much larger galaxies is fairly common (Fakhouri, Ma & Boylan-Kolchin 2010). The merger rate between dark matter haloes in the Millennium simulation is well fit by the functional form:

$$\frac{dN_m}{d\xi dz}(M, \xi, z) = AM^\alpha \xi^\beta \exp\left[\left(\frac{\xi}{\xi_{\text{crit}}}\right)^\gamma\right] (1+z)^\eta, \quad (3)$$

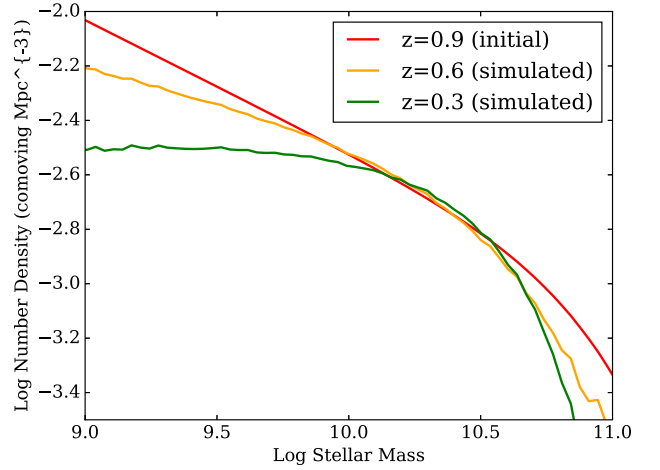
where  $N_m$  is the number density of mergers,  $M$  is the mass of the larger galaxy,  $\xi$  is the mass ratio between the smaller and larger galaxy,  $A$  is a normalization constant,  $z$  is redshift, and  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\eta$  are parameters determined by the physical model. The stellar mass–halo mass relation from Behroozi et al. (2013) was used to determine the effects of these mergers on stellar mass functions. As with star formation, the simulation calculates the number density of mergers and the resulting effect on the binned mass function in small time steps, stepping through the desired redshift range.

Small galaxies are indeed absorbed in mergers far more often than their more massive counterparts. Thus, in the absence of star formation, mergers alone will result in a shallower low-mass mass function slope towards lower redshift (Fig. 4). This acts in the opposite direction of the steepening due to evolution along the star-forming main sequence.

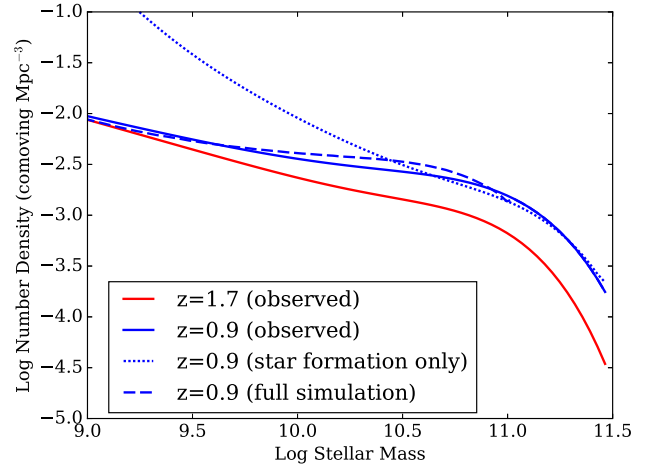
#### 4.1 Matching observed mass functions

Because dark matter simulations provide insufficient constraints on merger rates for our purposes, instead we adopt a similar approach to our analysis of quiescent populations. The galaxy population is divided into star-forming and quiescent populations, with quiescent populations constrained to match observed quiescent mass functions. Then, a merger rate function is selected, the galaxy population is evolved forward in time and the final mass function compared with observations.

Using the merger rate functional form from Fakhouri et al. (2010) with the proper slope and normalization parameters chosen from Fig. 6 to evolve galaxies from  $z = 1.7$  to  $z = 0.9$  produces a promising result (Fig. 5). Fitting the  $z = 0.9$  simulated



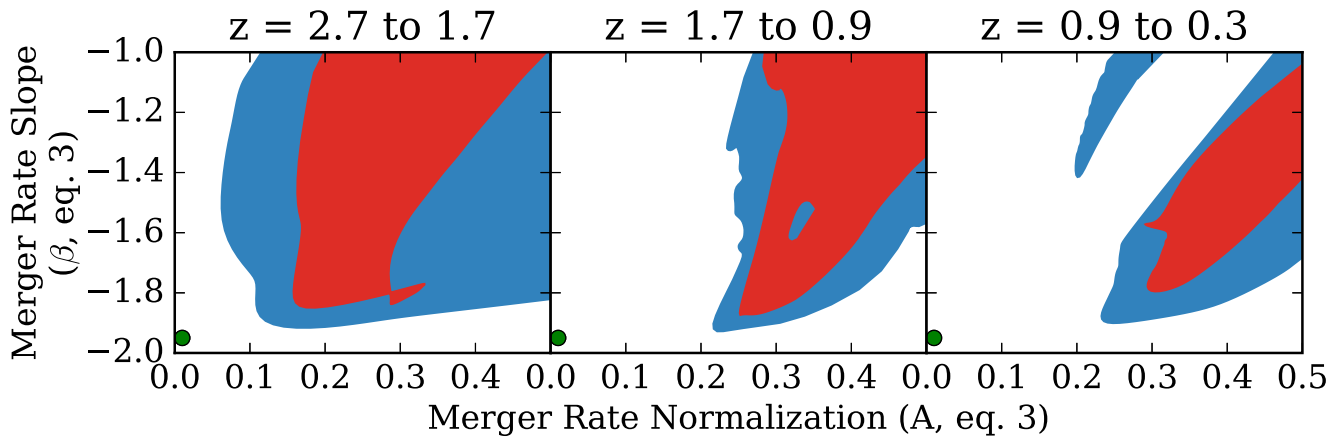
**Figure 4.** Simulation considering the effects of mergers in isolation, beginning with the observed  $z = 0.9$  mass function and evolving forward to  $z = 0.3$ . The low-mass slope becomes flatter, potentially counteracting the effects of star formation. Although it may appear mass is not conserved, the simulation finds that these seemingly ‘missing’ galaxies end up in massive clusters, well beyond the boundaries of this figure.



**Figure 5.** Simulation of galaxies evolving only with main-sequence star formation (dotted), and a combination of main-sequence star formation, quiescence and mergers (dashed) from  $z = 1.7$  to  $z = 0.9$ . This uses total mass functions from Ilbert et al. (2013) and the merger rate function form from Fakhouri et al. (2010) with well-chosen parameters. In both evolution cases parameters were picked to match the total mass function at the high-mass end ( $\log M > 11$ ). With star formation only we get far too many low-mass galaxies, but with mergers these are absorbed out of the population and we get a close match with observations.

curve to the double Schechter form from Ilbert et al. (2013) at  $z = 0.8$ – $1.1$  with a fixed  $\mathcal{M}^* = 10.87$  yields  $(\Phi_1^*, \alpha_1, \Phi_2^*, \alpha_2) = (2.09, -0.57, 0.39, -1.53)$ . This corresponds to an average  $\chi^2$  of 0.42 across the four parameters, which indicates possible consistency although it is difficult to properly account for the degrees of freedom allowed by our model, as the parameters are partially degenerate.

The merger rate function produced by simulations depends upon complex physics, and varying those physical parameters is beyond the scope of this paper. Instead, we elect to describe possible merger rate functions in terms of the functional form given in (3), producing



**Figure 6.** Choices of normalization ( $A$ , equation 3) and slope ( $\beta$ , equation 3) for merger rate functions producing mass functions consistent with observations at the  $1\sigma$  (red) and  $2\sigma$  (blue) levels for simulations run over three redshift ranges:  $2.7 < z < 1.7$  (left),  $1.7 < z < 0.9$  (centre) and  $0.9 < z < 0.3$  (right). There is some overlap between these regions, so a common merger rate functions would suffice over all three redshift ranges. The parameters corresponding to Fakhouri et al. (2010) are shown on each panel in green.

two parameters: a slope and normalization. A grid of possibilities is then evaluated to determine which are consistent with observations.

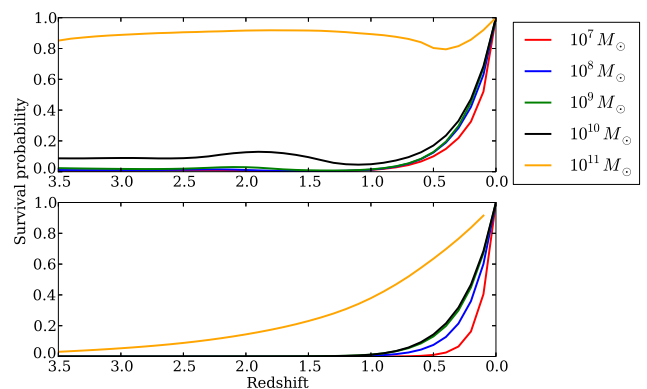
Our simulation is run over three redshift ranges:  $2.7 < z < 1.7$ ,  $1.7 < z < 0.9$  and  $0.9 < z < 0.3$ . In each case, some parameters are indeed consistent with the observed mass function at the final redshift (Fig. 6). The complex shape of the allowed parameter space is primarily due to differences on the high-mass end resulting from the observed mass function having been approximated with a Schechter function at both the initial and final redshifts. For some choices of parameters, similar merger rate functions are consistent with observations over all three redshift ranges. The merger rate function given by Fakhouri et al. (2010) does not lie in this allowed space. However, similar merger rate functions would be plausible.

#### 4.2 Comparing observations with quantitative merger simulations

It should be noted that the merger rate functions produced here describe luminous galaxies rather than the dark matter haloes described by numerical simulations of hierarchical merging. Thus, although it is tempting to interpret the slope and magnitude of the merger rate functions in Fig. 5 as candidates for dark matter halo merger rates, there are several reasons to be wary of a direct comparison. A proper translation requires not just matching halo masses with stellar masses, but also understanding how halo occupation rates vary between haloes of the same mass in rich environments with frequent mergers and in sparse environments or even voids.

If all haloes have a constant occupation rate, independent both of mass and of the likelihood that they will merge, then this effect is negligible. For example, if 50 per cent of haloes have galaxies, there are twice as many haloes as calculated from galaxy mass function at every mass, and thus four times as many mergers between any pair of masses. However, only 1/4 of those mergers will involve occupied haloes, producing the same rate for merging a pair of galaxies at those masses.

However, it is very likely that haloes in dense environments are more likely to be occupied, and thus the translation between halo merger rates and their effect on galaxy mass functions is more complex. In that regard, the merger rate functions described here are degenerate combinations of halo merger rates and occupation fraction. They describe merger rates between luminous galaxies,



**Figure 7.** The probability that (top) star-forming and (bottom) quiescent galaxies will survive (not be absorbed by a larger galaxy via merger) as a function of mass and redshift, according to prescription given in Fakhouri et al. (2010). Note that these probabilities are dominated by rare, high-mass galaxies, making them difficult to constrain.

and care must be taken when comparing them with predicted merger rates for dark matter haloes.

The rate at which a relatively small galaxy ( $10^9$  solar masses in Fig. 7) is absorbed into larger galaxies of various masses is strongly mass dependent (equation 3). Given current mass functions, we find that most small galaxies involved in mergers are absorbed by large ( $\log M_* > 11$ ) galaxies, even though there are relatively few large galaxies in the population. As a result, the overall merger rate of small galaxies is very sensitive to the total number of large galaxies.

Surprisingly, even relatively high-mass galaxies are likely to be absorbed by a larger object. For example, a  $M_* = 10^{10} M_\odot$  star-forming galaxy at redshift 3.5, despite being one of the largest galaxies in existence at that time, has only a 9 per cent probability of not merging with a larger object by redshift 0. For these galaxies, the larger object is likely the formation of a massive galaxy cluster, because individual galaxies are not seen with, e.g.  $M_* \sim 10^{14} M_\odot$ . Thus, these galaxies will often indeed survive to redshift 0, although many studies of mass functions exclude them and focus on field galaxies (cf. Bundy et al. 2015).

The uncertainty in the observed galactic mass function is relatively large at the high mass end due to the rarity of high-mass galaxies. There is a corresponding increase in uncertainty at the high-mass end of the stellar mass–halo mass relation (particularly at high redshifts), making it hard to match high stellar mass galaxies to their simulated halo merger rates. Varying these parameters does not significantly affect the overall shape of the merger rate function, so the flattening effect that mergers have on the overall mass function is fairly robust qualitatively. However, the slope and magnitude of this merger rate curve, and therefore the importance of mergers in determining mass functions, cannot be well constrained from halo simulations.

## 5 DISCUSSION

The star-forming main sequence appears to exist at all redshifts where it can be measured. However, its tightness and universality would appear to disagree with both theoretical expectations and observational evidence that environment and merger histories are also key drivers of galaxy evolution. Further, a continuity analysis shows that growth along the main sequence alone cannot reproduce observed stellar mass functions. In this work, we propose a possible solution: the star-forming main sequence is indeed a good description of star formation, but merging is also necessary to describe the growth of stellar mass. Counterintuitively, postulating a complete separation between the process of acquiring the ingredients that will become stars (clustering and merging) and the process of turning that into new stars (main-sequence star formation) can produce a model consistent with both theoretical and observational constraints.

We also note that a consistent model can be developed in which the duty cycle for star formation is 100 per cent. Galaxies would spend one long, extended period forming new stars along the main sequence followed by permanent quiescence. This is consistent with predictions made via measurements of synchronization timescales along the main sequence and with quasars (Steinhardt & Speagle 2014).

The next step is to understand whether this empirical model can be described in terms of fundamental astrophysical processes. Merging is responsible for the acquisition of hot gas, yet star formation requires further cooling (cf. Bromm 2013). Perhaps, the rate-limiting step lies in cooling channels, and as long as there is a surplus of hot gas available, star formation will be similar whether that hot gas was added in a recent merger or in the distant past. In that case, the time dependence of the star-forming main sequence may be due to declining background galaxy temperatures, perhaps associated with a cooling cosmic microwave background. Strong correlations between SFR and  $H_{II}$  regions yet weaker correlations with  $H_{I}$  (cf. Bigiel et al. 2008) may hint at a similar outcome.

This model also presents requirements for the merger rate function, reducing a two-dimensional space of possible parameters to a one-dimensional locus. Current simulations are underconstrained at high-redshift, so this produces a useful new set of constraints.

These constraints are primarily produced through a focus on the slope of the low-mass end of the observed stellar mass function. Several additional effects become important on the high-mass end, most notably stellar stripping (Murante et al. 2007; Purcell, Bullock & Zentner 2007; Guo et al. 2012; Contini et al. 2014). Properly modelling stellar stripping requires tracking additional history and is dependent upon environment, while one of the surprising features of the model in this paper is that the low-mass end can be matched with a history-free model, treating the various pro-

cesses involved as independent and random. Since stellar stripping is negligible below the ‘knee’ of the stellar mass function (Contini et al. 2017), it potentially provides a mechanism for improving the fit between predicted and observed mass functions through independently correcting for differences on the high-mass end.

Finally, we note that this model potentially provides a mechanism for explaining another longstanding puzzle, one of several effects often termed downsizing (Cowie et al. 1996). It is observed that typical massive galaxies finish their formation earlier than lower mass galaxies. However, smaller haloes should form earlier (Sheth et al. 2001; Springel et al. 2005; Vogelsberger et al. 2014), and the main sequence similarly indicates that low-mass galaxies form more efficiently. Considerable work has therefore gone into searching for models in which early, massive galaxies might become more efficient (e.g. Somerville & Davé 2015). Our new model proposes an entirely different solution: low-mass galaxies indeed form efficiently at high redshift, but they have a high probability of being absorbed into more massive ones.

The new empirical model presented in this work suggests several promising possibilities for unifying several different processes involved in galaxy evolution. If these behaviours can be produced via reasonable astrophysics, it could provide a solution to several key problems, most notably the remarkable disconnect between the star-forming main sequence and the remainder of our knowledge about how galaxies grow. However, this requires that such a model can be generated from meaningful cosmology and astrophysics, which requires considerable further analysis.

## ACKNOWLEDGEMENTS

The authors would like to thank Micaela Bagley, Andrew Benson, Zachary Claytor, Andreas Faisst, Olivier Ilbert, Adam Jermyn, Ian Kuehne, Dan Masters, Daniel McAndrew and Sune Toft for helpful comments. CLS acknowledges support from the ERC Consolidator Grant funding scheme (project ConText, grant No. 648179) and from the Carlsberg Foundation. DY was supported by the Victor Neher SURF Fellowship.

## REFERENCES

- Abell G. O., Corwin H. G. Jr, Olowin R. P., 1989, *ApJS*, 70, 1
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2846
- Bouwens R. J. et al., 2015, *ApJ*, 803, 34
- Bromm V., 2013, *Rep. Prog. Phys.*, 76, 112901
- Bundy K. et al., 2015, *ApJS*, 221, 15
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Contini E., De Lucia G., Villalobos Á., Borgani S., 2014, *MNRAS*, 437, 3787
- Contini E., Kang X., Romeo A. D., Xia Q., 2017, *ApJ*, 837, 27
- Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, *AJ*, 112, 839
- Daddi E. et al., 2007, *ApJ*, 670, 156
- Duncan K. et al., 2014, *MNRAS*, 444, 2960
- Elbaz D. et al., 2007, *A&A*, 468, 33
- Fakhouri O., Ma C.-P., 2009, *MNRAS*, 394, 1825
- Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, *MNRAS*, 406, 2267
- Genel S., 2016, *ApJ*, 822, 107
- Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, *ApJ*, 778, 14
- Guo Y. et al., 2012, *ApJ*, 749, 149
- Hartley W. G., Conselice C. J., Mortlock A., Foucaud S., Simpson C., 2015, *MNRAS*, 451, 1613
- Hayward C. C. et al., 2014, *MNRAS*, 445, 1598

- Hirschmann M., De Lucia G., Wilman D., Weinmann S., Iovino A., Cucciati O., Zibetti S., Villalobos Á., 2014, *MNRAS*, 444, 2938
- Ilbert O. et al., 2013, *A&A*, 556, A55
- Kawinwanichakij L. et al., 2016, *ApJ*, 817, 9
- Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531
- Khabiboulline E. T., Steinhardt C. L., Silverman J. D., Ellison S. L., Mendel J. T., Patton D. R., 2014, *ApJ*, 795, 62
- Leauthaud A. et al., 2012a, *ApJ*, 744, 159
- Leauthaud A. et al., 2012b, *ApJ*, 746, 95
- Lee N. et al., 2015, *ApJ*, 801, 80
- Leja J., van Dokkum P. G., Franx M., Whitaker K. E., 2015, *ApJ*, 798, 115
- Murante G., Giovalli M., Gerhard O., Arnaboldi M., Borgani S., Dolag K., 2007, *MNRAS*, 377, 2
- Noeske K. G. et al., 2007, *ApJ*, 660, L43
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Peng Y.-j., Lilly S. J., Renzini A., Carollo M., 2014, *ApJ*, 790, 95
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Purcell C. W., Bullock J. S., Zentner A. R., 2007, *ApJ*, 666, 20
- Satyapal S., Ellison S. L., McAlpine W., Hickox R. C., Patton D. R., Mendel J. T., 2014, *MNRAS*, 441, 1297
- Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, *ApJS*, 214, 15
- Springel V. et al., 2005, *Nature*, 435, 629
- Steinhardt C. L., Speagle J. S., 2014, *ApJ*, 796, 25
- Steinhardt C. L. et al., 2014, *ApJ*, 791, L25
- Szomoru A., van Gorkom J. H., Gregg M. D., Strauss M. A., 1996, *AJ*, 111, 2150
- Tasca L. A. M. et al., 2015, *A&A*, 581, A54
- Tomczak A. R. et al., 2014, *ApJ*, 783, 85
- Tomczak A. R. et al., 2016, *ApJ*, 817, 118
- van Dokkum P. G., 2005, *AJ*, 130, 2647
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, *MNRAS*, 372, 1161
- Whitaker K. E. et al., 2014, *ApJ*, 795, 104

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.